

Research Space

Journal article

**Digital mammographic interpretation by UK radiographer
mammographers: A JAFROC analysis of observer performance**
Williams, S., Aksoy, U., Reed, W., Cielecki, L. and Woznitza, N.

<https://doi.org/10.1016/j.radi.2021.02.015>

Introduction

Breast cancer is the most common cancer in females, accounting for 31% of all new female cancer cases in the UK in 2016 with approximately 11,400 deaths per year(1). Screening mammography is widely used in the UK as a method to detect earlier stage breast cancer, with women aged between 50 and 70 invited to routine breast screening every 3 years (2). A skilled mammography reader will be able to perceive and identify a potential abnormality such as a breast cancer from a wide range of mammographic features. As part of any cost effective screening programme an equally challenging skill is the ability to recognise and disregard benign findings (3,4). The mammography reader requires high level cognitive, perceptual and analytical skills to detect or exclude the presence of breast cancer (5).

Historically it was considered that this required a medical and specialist qualification in breast radiology (6) in order to make these complex conclusions. However, it has been demonstrated that suitably trained radiographer readers have comparable results to radiologists in screening and symptomatic settings (7,8). Whilst radiographer readers are well established in the UK the model is not as well established internationally.

An efficient method of measuring readers' performance, such as screen reader test sets is needed, to foreshorten the time taken to identify under-performance and instigate appropriate quality improvement programmes in a timely fashion. Clinical audit has been used with good effect to assess screen readers' performance, but does present certain limitations, which have encouraged the development of supplementary strategies. One such strategy is the provision of standardized mammographic screen reading test sets, like PERFORMS (Personal Performance in Mammographic Screening) implemented by the National Health Service Breast Screening Programme (NHS BSP) in the UK in 1991 (9) or BREAST (BreastScreen REader Assessment STRategy) in Australia. In Australia, BREAST has been used as a national quality training tool in BreastScreen services since 2011 providing readers with a variety of performance scores and immediate, individual feedback on missed cancers and false-positive selections (10).

The ability to accurately characterise mammographic features is essential to a sustainable breast screening program regardless of professional background. For many years radiographers have been involved in the reporting and interpretation of mammograms (7,8,11) and it is essential they perform to the expected radiology standards. This study using the BREAST test program aims to demonstrate the UK radiographer reader performances and analyse the variabilities of interpretive accuracies.

Methods

Study design

The purpose of the test was to establish diagnostic accuracy in interpreting mammograms. A prospective study of radiographers trained to review and interpret mammographic images was performed under test conditions. The reader performance data was collected at Symposium Mammographicum in 2018, a biannual conference in the UK (12). The study utilised the web-based system (BREAST) with an extensive database of full field digital

mammographic (FFDM) images with previously obtained ethical approval. The test set allocated by BREAST had been previously used in other studies (13,14). All of the cases had previously been validated and verified with pathology truth established. The test set comprised of 60 standard view, challenging cases; 42 with prior imaging for comparison and 18 with no previous imaging. The distribution of mammographic examinations was designed to resemble clinical prevalence, albeit with a higher number of abnormalities; 20 cases with biopsy proven malignancies and 40 normal cases, ordered randomly. Cases were confirmed by two experienced radiologists and follow-up negative screening mammograms obtained in the succeeding screening round. The 20 positive cases contained a variety of lesion sizes and malignancy appearances.

The images were pre-loaded onto two PACS workstations (Barco Coronis Uniti (MDMC)12MP display) meeting the diagnostic standards for reporting mammograms with a standard hanging protocol and running order. The participants recorded their findings directly onto the BREAST web-site on a separate laptop as illustrated in figure 1. The workstations were in a dedicated room with conditions appropriate for image assessment.



Figure 1. A diagrammatical representation of the workstation setup.

On completion, the results of the group were compared to previous studies for groups where a first BREAST test set had been recorded; some were part of an ongoing study where additional BREAST tests were completed. It was not possible to determine which participants, if any, had taken the Sydney BREAST test set and therefore direct comparison was not possible. For the purposes of this study an assumption was made that as a validated tool used for measuring performance all BREAST test sets are of an acceptable standard and comparison has validity.

Participants

Participants were self-selected from conference delegates. Any radiographers trained in reading and interpreting mammography images and currently practicing in the UK were eligible for the study and booked a designated session to complete the test set. Radiologists, radiographers training to be readers and any other type of reader such as clinicians were excluded.

Four participants worked in a diagnostic setting and six in the NHS BSP program. The mean number of years' experience in interpreting mammograms was 4.5 (range 1-14). Five participants were reading 5000 or more mammograms per year and five were reading less than 5000 per year.

Participants booked a 90 minute slot to complete the test set. Each participant was given a unique BREAST account with user name and password and was required to complete a short paper questionnaire about their mammography reading background. A participant consent form was presented to each individual at the start of their session, embedded in the programme software. Ethical approval was obtained from the University of Sydney for an international reader study. Informed and written consent was obtained from participants prior to data collection.

Process

Instructions were given both verbal and written, explaining how to view the images and record their decisions. A copy of the instructions was made available at each workstation for reference during each session.

A free-response methodology was used (15). Participants reviewed each case in turn and were able to identify any mammographic feature they considered relevant, including multiple features in the contralateral or ipsilateral breast. The mammographic feature was chosen from a pre-selected menu as shown in table 1.

Table 1 – Mammographic features

Mammographic feature
Calcifications
Stellate
Discrete mass
Spiculated mass
Non-specific density
Architectural distortion
Lymph node

They were required to localise any abnormality by marking them, in one or both projections, electronically on the corresponding images on the web-site. The participant marked each mammographic feature and provided a confidence rating for a cancer being present to give a mark rating pair. This was embedded in the BREAST program. A summary of the confidence ratings is shown in Table 2.

Table 2: Confidence level for breast cancer

Confidence rating	Decision
1	Normal
2	Benign
3	Likely to be cancer
4	Highly likely to be cancer
5	Malignant

Any case with no features marked for review was automatically recorded as normal.

During the test participants were able to move between cases and to manipulate images to enhance visualisation. Once all 60 cases were completed the participant submitted the results for analysis and could compare their opinion with the reference image as shown in figure 2.

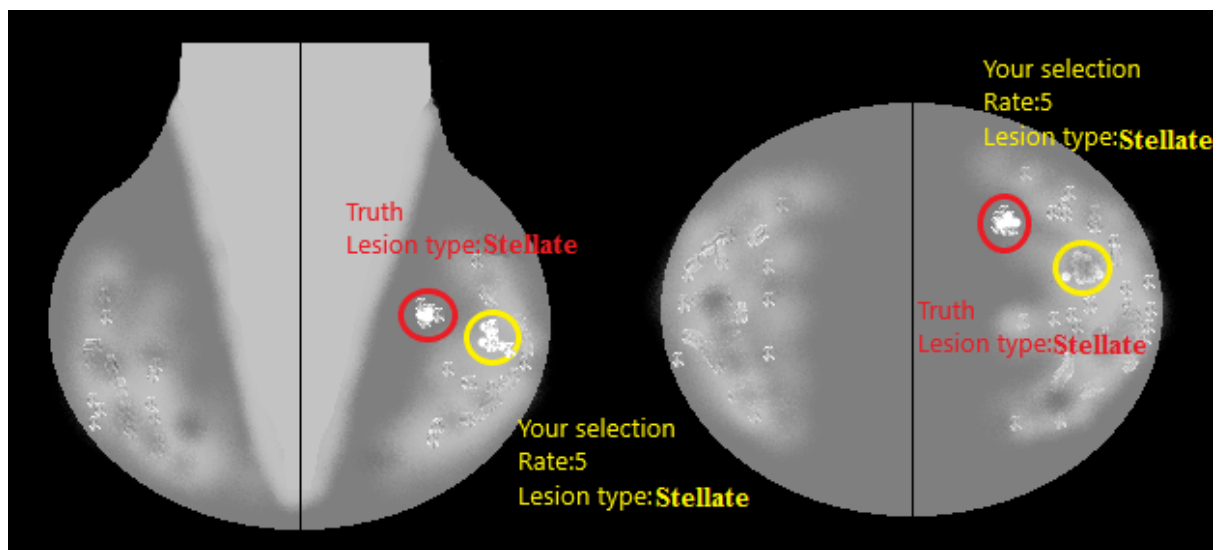


Figure 2. Diagrammatical representation of the pathology truth compared to a selected feature and confidence rating

Participants were measured against the correct identification of cancer (true positive; sensitivity) and normal (true negative; specificity) cases. A lesion incorrectly located in the breast on one view was considered correct for sensitivity but incorrect for location sensitivity.

Participants performance values included receiver operating characteristic (ROC) and jack-knife alternate free-response receiver operating characteristic (JAFROC) figures of merit, sensitivity, location sensitivity, specificity, true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). All data produced were de-identified and stored on a cloud server and downloaded onto a central database for analysis as part of the study.

Data Analysis

Mammogram cases could be correctly identified as abnormal but the actual cancer location incorrectly marked on the image; therefore analysis was performed to reflect both case-based (ROC score) and lesion-based accuracy (JAFROC score). The case-based analysis recorded a correctly identified abnormal mammogram; but did not reflect if the correct mammographic feature had been marked as the malignancy. The lesion-based analysis recorded how accurately the participant had marked on the images (location sensitivity) as defined by a 75-pixel radius. Each feature selected corresponded to a pixel reference on the X and Y axis which was cross referenced with the actual co-ordinates of the cancer. The analysing program allows for reduced image quality on the web-based images when the features are marked, for all participants and any test set. Each case was categorised as FN, TP, FP or TN and the ability of the reader to correctly identify the abnormality (sensitivity) and recognise normal/benign (specificity) was calculated for the test group. A comparison was made to previous studies that had undertaken the same or similar validated test sets for the first time (13,14,16).

Results

The results of our study have shown that under test conditions 10 trained radiographer readers were able to recognise normal features and identify and locate cancer on mammographic images. Six participants were currently active in breast screening. Five participants were consultant radiographers/mammographers; one trainee consultant radiographer; two advanced practitioners and a clinical specialist. Table 3 summarises the background of each of the participants.

Table 3 Background of the participants

Reader	Are you a qualified mammography image reader?	Do you work in Breast Screening?	How many years have you been reading mammography images?	What is your job title/professional background?	How many images do you read per year?
A	Y	Y	6	Consultant Radiographer	10000+
B	Y	Y	4	Consultant Mammographer	3500

C	Y	N	1	Advanced Practitioner	500
D	Y	N	4	Clinical Specialist	2000
E	Y	N	2	Consultant Mammographer	4000
F	Y	N	14	Consultant Radiographer	7000
G	Y	Y	2	Advanced Practitioner	5000
H	Y	Y	3	Advanced Practitioner and Superintendent	15000
I	Y	Y	8	Consultant Radiographer	5000
J	Y	Y	1	Trainee Consultant Radiographer	3500
			Mean: 4.5 (range 1-14) years		

The comparison to similar studies is summarised in table 4. In each of these studies the participants completed a test set of similar difficulty under test conditions. The study group scored 83 for sensitivity (95% confidence interval 72.5%-93.6%; range 66 - 85) and above the mean of 79.46. The study group scored 69.3 for specificity (95% CI 52.6%-85.9%; range 63.9-85) and below the mean of 73.6. The mean for lesion sensitivity was 46.5 with the test group scoring 74.8 (95%CI 64.2%-85.4%; range 32.5 – 74.8) thus scoring the highest of the test groups that measured lesion sensitivity. Mean figure of merit and area under the curve for the study cohort were 0.74 (95%CI 0.933-0.779) and 0.86 (95%CI 0.824-0.886).

Table 4 Assuming all BREAST test sets are of a similar standard - test results and comparison with other studies for first test set taken by participants of each study group

	Our study (n=10)	Trieu et al (2019) (12) (n=40)	Suleiman et al (2016) (13) (n=14)	Trieu et al (2019) (12) (n=17)	Trieu et al registrars(2019) (12) (n=10)	Soh et al (2016) N=53 (Aus)	Soh et al (2016) N=15 (Sing)
Sensitivity	83	84.4	74	83.8	66	85	80
Specificity	69.3	75.2	67	74.8	63.9	80	85
Lesion sensitivity	74.8	64	51	60.7	32.5	-	-
ROC	0.86	0.85	0.79	0.85	0.66	0.86	0.86
JAFROC	0.74	0.76	0.61	0.74	0.43	0.80	0.72

Discussion

BREAST test sets are designed to challenge the individual undertaking the test. In Australia they are recognised as both a way of measuring radiologist performance and as an established learning tool (10). However identifying how a case may be classified into levels of difficulty is complicated and multifaceted (17) and test results may not translate into clinical practice. Participant performance for PERFORMS and BREAST test sets have been shown to be a strong indicator for translating to clinical performance (18,19).

The BREAST test set could offer an insight into performance under similar conditions even with inherent limitations (20). All test sets for BREAST are of an equivalent standard and have been used in studies (13,14) to measure radiologist performance establishing acceptable statistical standards against which other readers can be measured. The tests sets have been used to make comparisons with radiologists from other countries and found comparable performance standards (16). These standards were reached by the participants of our study when compared to similar studies of radiologists for sensitivity, specificity and lesion sensitivity. In addition, our radiographer reader study group performed highest of all the compared groups that measured lesion sensitivity. These findings suggest that factors other than background profession can have a significant influence on the performance of the individual under test conditions. Studies have suggested both personal and external factors influence performance and will cause performance variation including, professional support networks, workload, experience and, education (21-23). Our study suggests the correct clinical evaluation of mammograms is not necessarily role dependent but task and or experience dependent. Previous studies of radiographer performance in reading mammograms has shown that exposure to mammograms as part of their experiential learning in breast imaging may give them a level skill which enable them to recognise mammographic abnormalities on standard mammograms. Studies in the Netherland found that the initial and on-going training of mammography technologists (mammographers) showed a high prevalence of breast cancer in cases they deemed to be abnormal (24-26).

Suleiman et al (2016) suggested that structured educational strategies could be used to improve reader performance. A recent Australian study of experienced mammographers when compared to other similar UK radiographer reader studies, (11) suggested performance differences to be likely the result of dedicated and extensive education in mammogram interpretation (27). Any reader following such a structured programme could be reasonably compared to and referenced against, the established standards. In Mexico after following the same 6 months training regime as a radiologist, radiologic technologists (mammographers) had comparable results under test conditions to a radiologist (28). In considering education strategies Scott and Gale (2006) explored variation between occupational groups in using structured test sets such as PERFORMS to target learning needs based on occupational groups. They found no significant performance difference between radiographers and radiologists when matched for other varying factors, after 3 sets of 60 cases (29), thus supporting the findings of our study.

An important aspect of any test set is the relevance of this in translation to everyday clinical practice. Normal clinical practice for breast screening conditions yields a relatively low number of breast cancers within the screened population. Under the scrutiny of test conditions and expectations of the test the participant expects to find a relatively high number of abnormalities, however, learning to do the test set itself, as learning to do the test may have significant influence on performance. The test set up and equipment was unfamiliar to the participants as well as the challenge of the test. To compensate for this our study has made comparison to other studies where it was the first time the radiologists had taken a BREAST test set (13,14,16) which may have been part of an ongoing study where additional BREAST tests were completed. As test settings and taking the test itself has influence on behaviour further research is required to explore if the findings of our study will translate into the clinical setting. A recent study by Chen et al (2020) demonstrated a potentially predictive correlation between PERFORMS test and clinical performance of individual readers in the UK regardless of their role and further study is warranted (30). The study group showed variation in their clinical settings and experience; studies to explore the influences of these factors are warranted to determine influence these factors have on performance.

The chronic workforce issues for breast imaging services in the UK (31) and in other countries have driven studies (16,24-28) exploring the utilisation of radiographers in the timely and safe diagnosis of breast cancer. Our study has shown that this is an option that should be given more consideration for radiographers internationally.

A limitation identified of this small group study is that all participants were from the same occupational group attending a single professional conference. A larger group of participants would verify our test results which would allow subgroup analysis of radiologists, radiographers and other groups. The effect of different educational backgrounds and clinical settings could also be evaluated. Another limitation is as the task was performed under specific test conditions the results may not necessarily translate into clinical practice.

Conclusion

Under test conditions UK radiographer readers demonstrated a performance comparable to international radiologists using a BREAST test set. Further study is required with a larger cohort to explore if this would translate to a wider population of radiographer readers.

Implications for clinical practice

Severe workforce issues in breast imaging in the UK and worldwide mean different models of service delivery need to be considered to provide sustainable safe breast services. Our study suggests translation of test findings into clinical practice will help address access and capacity issues in the timely identification and diagnosis of breast cancer.

References **References**

1. Cancer research UK. Breast cancer incidence (invasive) statistics. Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/incidence-invasive#ref-0> [Accessed 22 Nov 2019]
2. Breast screening programme England 2018-19. Screening and Immunisation team 2020. Available from: <https://files.digital.nhs.uk/0A/9D9F34/breast-screening> [Accessed 28 May 2020]
3. Mittmann N, Stout NK, Lee P, Tosteson ANA, Trentham-Dietz A, Alagoz O et al. Total Cost-effectiveness of Mammography Screening Strategies. *Health Rep.* 2015;**26**(12): 16-25.
4. Pharoah PDP, Sewell B, Fitzsimmons D, Bennett HS and Pashayan N. Cost effectiveness of the NHS breast screening programme: life table model. *BMJ* 2013 <http://doi.org/10.1136/bmj.f2618>
5. Carrigan AJ, Wardle SG, Rich AN. Finding cancer in mammograms: if you know it's there, do you know where? *Cognitive Research: Principles and Implications* 2018 **3**:10 <http://doi.org/10.1186/s41235-018-0096-5>
6. Donovan T and Manning DJ. Successful reporting by non-medical practitioners such as radiographers, will always be task-specific and limited in scope. *Radiography* 2006;**12**(1):7-12 <http://doi.org/10.1016/j.radi.2005.01.04>
7. Culpan AM. Radiographer involvement in mammography image interpretation: A survey of United Kingdom practice. *Radiography* 2016;**22**:306-312 <http://doi.org/10.1016/j.radi.2016.03.004>
8. Bennett RL, Sellars SJ, Blanks RG, and Moss SM. An observational study to evaluate the performance of units using two radiographers to read screening mammograms. *Clin. Radiol.* 2012 **67**(2):114-21 <http://doi.org/10.1016/j.crad.2011.06.015>
9. Gale A. PERFORMS: a self-assessment scheme for radiologists in breast screening. *Semin Breast Dis* 2003;**6**:148-52. <http://doi.org/10.1053/j.sembd.2004.03.006>
10. BREAST. BreastScreen reader assessment strategy. Available from: <http://www.breastaustralia.com/> [Accessed 20 May 2020].
11. Wivell G, Denton ERE, Eve CB, Inglis JC, Harvey I. Can radiographers Read screening Mammograms? *Clin. Radiol.* 2003;**58**(11): 63-67 <http://doi.org/10.1016/j.crad.2003.08.002>
12. Symposium mammographicum. Available from: <https://sympmamm.org.uk/> [Accessed 4 Dec 19]
13. Trieu PD, Tapia K, Frazer H, Lee W; Brennan P. Improvement of Cancer Detection on Mammograms via BREAST Test Sets. *Acad Radiol* 2019;**26**: e341e347 <http://doi.org/10.1016/j.acra.2018.12.017>
14. Suleiman WI, Rawashdeh MA, Lewis SJ, McEntee MF, Lee W, Tapia K, et al. Impact of Breast Reader Assessment Strategy on radiologists' test performance. *Journal of Medical Imaging and radiation Oncology* 2016;**60**(3): 352-358 <http://doi.org/10.1111/1754-9485.12461>
15. Chakraborty DP. Recent advances in observer performance methodology: jackknife free-response ROC (JAFROC). *Radiation protection dosimetry* 2005;**114**(1-3):26-31. <http://doi.org/10.1093/rpd/nch512>
16. Soh BLP, Lee WB, Wong J, Sim L, Hills SL, Tapis KA et al. Varying performance in mammographic interpretation across two countries: Do results indicate reader or population variance? *Medical Imaging 2016: Image perception, observer*

Performance and Technology Assessment. 2016;Vol 97870X

<https://doi.org/10.1117/12.2217654>

17. Norsuddin NM, Mello-Thoms C, Reed W, Rickard M, Lewis S. An investigation into the mammographic appearances of missed breast cancers when recall rates are reduced *Br J Radiol* 2017;**90**:20170048 <http://doi.org/10.1259/bjr.20170048>
18. Soh BP, Lee W, McEntee MF, Kench PL, Reed WM, Heard R, et al. Screening Mammography: Test Set Data Can Reasonably Describe Actual Clinical Reporting *Radiology*:2013; **268**(1):46-53 <http://doi.org/10.1148/radiol.13122399>
19. Scott HJ, Evans A, Gale AG, Murphy A, Reed J. The relationship between real life breast screening and an annual self assessment scheme. *Medical Imaging 2009: Image Perception , Observer Performance and Technology assessment* 2009. <http://spie.org/x33859.xml>
20. Soh BP, Lee W, Kench PL, Reed WM, McEntee, Poulos A, et al. Assessing reader performance in radiology, an imperfect science: lessons from breast screening. *Clinical Radiology* 2012;**67**:623-628 <https://doi.org/10.1016/j.crad.2012.02.07>
21. Tavakoli Taba S, Hossain L, Heard R, Brennan P, Lee W, Lewis S. Personal and Network Dynamics in Performance of Knowledge Workers: A Study of Australian Breast Radiologists. *PLOS ONE* 2016;**11**(2): e0150186. <https://doi.org/10.1371/journal.pone.0150186>
22. Rawashdeh MA, Lee WB, Bourne RM, Ryan EA, Pietrzyk MW, Reed WM, et al. Markers of good performance in mammography depend on number of annual readings. *Radiology* 2013;**269**: 61–7. <https://doi.org/10.1148/radiol.13122581>
23. Geller BM, Bogart A, Carney PA, Sickles EA, Smith R, Monsees B, et al. Educational Interventions to Improve Screening Mammography Interpretation: A Randomized Controlled Trial *AJR* 2014;**202**: W586-W596. <https://10.2214/AJR.13.11147>
24. Duijm L, Groenewoud J, Fracheboud J, de Koning H. Additional double reading of screening mammograms by radiologic technologists: impact on screening performance parameters. *J Natl Cancer Inst* 2007;**99**(15): 1162–70. <https://doi.org/10.1093/jnci/djm050>
25. Duijm L, Groenewoud J, Fracheboud J, van Ineveld B, Roumen R, de Koning H. Introduction of additional double reading of mammograms by radiographers: effects on a biennial screening programme outcome. *Eur J Cancer* 2008;**44**(9): 1223–8. <https://doi.org/10.1016/j.ejca.2008.03.003>
26. Duijm L, Louwman M, Groenewoud J, van de Poll-Franse L, Fracheboud J, Coebergh J. Inter-observer variability in mammography screening and effect of type and number of readers on screening outcome. *Br J Cancer* 2009;**100**: 901–7. <https://doi.org/10.1038/sj.bjc.6604954>
27. Moran S and Warren-Forward H. Can Australian radiographers assess screening mammograms accurately? First stage results from a four year prospective study. *Radiography* 2016;**22** e106-e111 <https://doi.org/10.1016/j.radi.2015.11.005>
28. Torres-Mejía, Smith RA, de la Luz Cararanza-Flores M, Bogart A, Martínez-Matsushita L, Miglioretti DL, et al. Radiographers supporting radiologists in the interpretation of screening mammography: a viable strategy to meet the shortage in numbers of radiologists. *BMC Cancer* 2015;**15**:410 <https://doi.org/10.1186/s12885-015-1399-2>
29. Scott HJ and Gale AG. Breast screening: PERFORMS identifies key mammographic training needs. *Br J Radiol* 2006;**79**:S127-S133. <https://doi.org/10.1259/bjr/25049149>

30. Chen Y, James JJ, Cornford EJ, Jenkins J. The Relationship between Mammography Readers' Real-Life Performance and Performance in a Test Set–based Assessment Scheme in a National Breast Screening Program. *Radiology: Imaging cancer* 2020;**2**(5) <https://doi.org/10.1148/rycan.2020200016>
31. The Royal College of Radiologists. The breast imaging and diagnostic workforce in the United Kingdom. Results of a survey of NHS Breast Screening Programme units and radiology departments. 2016. Available from: <https://www.rcr.ac.uk/publication/breast-imaging-and-diagnostic-workforce-united-kingdom>. [Accessed 27 Jan 2021]